

PENGUMPULAN DATA DAN PREDIKSI MASUK DI SEMUA SMP NEGERI KOTA CIMAHI MENGGUNAKAN METODE RANDOM FOREST

COLLECT DATA AND PREDICTION OF SIGNS JUNIOR HIGH SCHOOLS IN CIMAHI CITY BY USING RANDOM FOREST METHOD

Gerry Pratama¹, Muhammad Nasrun S.Si., M.T.², Anton Siswo Raharjo Ansori, S.T., M.T.³

^{1,2,3}Prodi S1 Teknik Komputer, Fakultas Teknik Elektro, Universitas Telkom

¹gerrypratamaaa@gmail.com, ²nasrun@telkomuniversity.ac.id, ³raharjo@telkomuniversity.ac.id

ABSTRAK

Tahun ajaran baru menjadi momen yang paling ramai dalam dunia pendidikan khususnya di Indonesia. Di setiap sekolah pasti memiliki beberapa persyaratan untuk masuk sekolah yang mereka inginkan, sebagai contohnya seperti dilihat dari nilai akhir ujian nasional dan juga dilihat dari jarak alamat rumah calon peserta didik dengan sekolah yg dituju atau biasa disebut juga zonasi, dan jalur prestasi tersebut menjadi persyaratan masuk sekolah di setiap SMP dikota Cimahi. Pada Tugas Akhir ini, Penulis mengusulkan mengumpulkan data dan memprediksinya menggunakan Metode Random Forest dan menerapkan metode tersebut kedalam sistem android *Smartphone* menggunakan android studio guna mempermudah agar para orang tua maupun siswa mampu mengakses dan mendapat persentase prediksi masuk sekolah yang dituju, dengan cara langsung memasang perangkat lunak tersebut di *Smartphone*, siswa maupun orang tua siswa lebih cepat menentukan pilihannya dan lebih praktis untuk kedepannya untuk memilih sekolah yang tepat.

Kata Kunci: Prediksi, Penerimaan Siswa, Random Forest

ABSTRACT

New academic year have become the most exciting moment in education, especially in Indonesia. Every schools definitely have some requirements to get into the school that they want to, for example by looking at the final result of National Examination and also looking at the distance of the prospective students' home addresses, zonation system. Previous school, and achievements path have become one of requirements to enter every Junior High School in Cimahi. In this final project, the author proposes, collect the data, and predict it by using Random Forest Method, and apply the method into android smartphone system using Android Studio in order to make parents and also students can access and get the prediction percentage of intended school. By directly installing the software on a smartphone, students and parents can make their choice faster and more practical for the future to choose the right school.

Keywords: Prediction, Student Admission, Random Forest

1. Pendahuluan

1.1 Latar Belakang

Saat ini tingkat keinginan masyarakat untuk bersekolah semakin meningkat. Namun di setiap sekolah pasti memiliki beberapa persyaratan untuk masuk sekolah yang mereka inginkan akan tetapi belum dapat diprediksikan informasi pentingnya, sebagai contohnya seperti dilihat dari nilai akhir ujian nasional dan juga dilihat dari jarak alamat rumah calon peserta didik dengan sekolah yg dituju atau biasa disebut juga zonasi, dan jalur prestasi tersebut menjadi persyaratan masuk sekolah di setiap SMP dikota Cimahi. Sehingga dibutuhkan sebuah sistem yang mampu memprediksi untuk memudahkan para orang tua yang ingin memasukan anaknya ke SMP Negeri yang diinginkan di Cimahi.

Penulis mengusulkan mengumpulkan data dan memprediksi masuk SMP Negeri di kota Cimahi menggunakan Random Forest (RF). Metode yang memprediksi sekolah ini juga ditunjang persentasenya dengan di tambah beberapa

faktor pendukung calon siswa dalam menginputkan data diantara lain seperti zonasi, dan jalur prestasi yang diharapkan akan mengeluarkan model yang dapat memprediksi nilai masuk sekolah untuk membantu para orang tua siswa yang ingin mendaftarkan anaknya masuk sekolah yang diinginkan berdasarkan nilai ujian nasional siswa tersebut.

1.2 Tujuan

Tujuan yang ingin dicapai dari pembuatan tugas akhir ini adalah:

1. Mengimplementasikan metode Random Forest tersebut dalam prediksi masuk SMP Negeri Di Kota Cimahi berbasis android.
2. Melakukan analisis terhadap hasil implementasi yang diperoleh.
3. Mengkaji keakurasian dari metode *Random Fosrest* dari input yang diberikan calon siswa yaitu nilai Ujian Nasional.
4. Menampilkan persentase prediksi masuk sekolah yang dituju.

1.3 Identifikasi Masalah

Rumusan masalah dalam tugas akhir ini adalah:

1. Bagaimana pengelompokan data dan cara mengolahnya sehingga data dapat di implementasikan kedalam android studio.
2. Membandingkan hasil dari sistem peramalan menggunakan metode random forest dengan data aktual untuk kemudian diperoleh tingkat keakuratan sistem.
3. Membuat persentase prediksi masuk sekolah dengan mengikuti aturan-aturan yang sudah ditetapkan dinas Pendidikan seperti ketentuan zonasi, Nilai Ujian Akhir siswa dan jalur prestasi dengan menggunakan metode Random Forest.

1.4 Metode Penelitian

a. Studi literatur

Studi literatur bertujuan untuk mempelajari dasar-dasar dari laporan tugas akhir, yang terdiri dari latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi penelitian, serta sistematika penulisan tugas akhir yang digunakan.

b. Landasan Teori

Perancangan sistem bertujuan menjelaskan tentang beberapa teori penunjang yang digunakan dalam tugas akhir meliputi penjelasan teori mengenai pendaftaran siswa baru, data mining, dan menjelaskan metode Random Forest beserta kinerjanya.

c. Analisis Dan Perancangan

Bertujuan untuk membahas hasil analisa dan perancangan yang meliputi pembahasan mengenai deskripsi metode, analisa perangkat lunak dan perancangan perangkat lunak meliputi *library* yang diunakan. Pengujian sistem bertujuan melakukan implementasi metode pada aplikasi sesuai dengan perancangan yang telah dilakukan.

d. Implementasi dan pengujian

bertujuan untuk membahas tentang implementasi pengujian, deskripsi pengujian, hasil pengujian berdasarkan beberapa percobaan yang telah dilakukan sehingga menghasilkan data yang tepat dan akurat.

2. Dasar Teori

2.1 Pendaftaran Siswa Baru

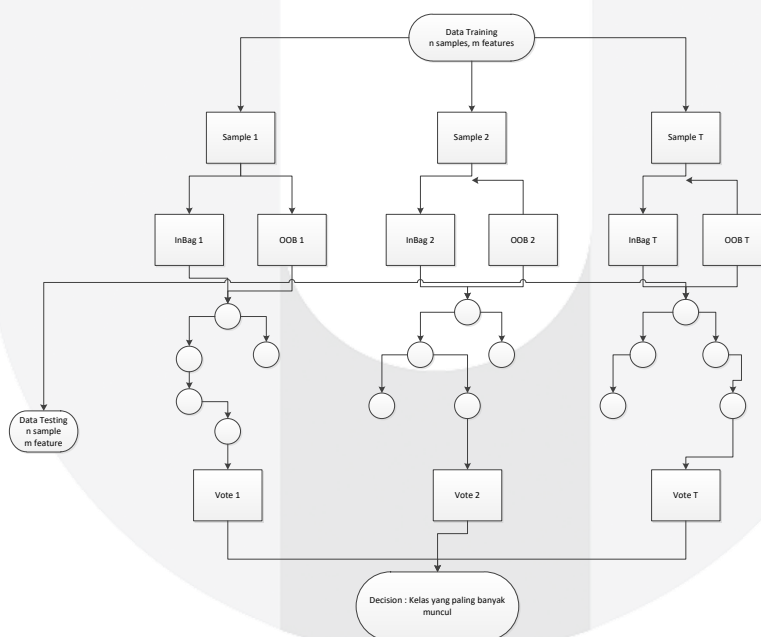
Tahun ajaran baru menjadi momen yang paling ramai dalam dunia pendidikan khususnya di Indonesia karena pada tahun ajaran baru setiap sekolah melakukan penerimaan siswa baru. Salah satu tujuan penyelenggaraan penerimaan siswa baru yaitu menjaring calon siswa baru yang memiliki kualifikasi yang sesuai dengan ketentuan dan persyaratan penerimaan siswa baru di Sekolah yang dituju. Pemerintah pusat mewajibkan semua sekolah untuk menerapkan sistem zonasi dalam menggelar Penerimaan Peserta Didik Baru (PPDB) tahun ajaran 2017/2018. Sistem zonasi sudah dirancang sejak tahun lalu, sekolah wajib menerima semua calon siswa yang berdomisili masih satu wilayah dengan sekolah. PPDB dengan sistem zonasi sesuai dengan Permendikbud nomor 17/2017 tentang PPDB. Penentuan zona diserahkan kepada sekolah dan pemerintah daerah. Pasalnya, kepadatan penduduk di setiap wilayah sangat beragam. Semua sekolah serius menerapkan sistem zonasi agar kualitas pendidikan nasional semakin baik tanpa diskriminasi.

2.2 Data Mining

Data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar. Data mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data.[13]

2.3 Random Forest

Random Forest (RF) adalah metode klasifikasi dan regresi berdasarkan agregasi sejumlah pohon keputusan. Secara khusus, merupakan sebuah *ensemble* (kumpulan) dari pohon yang dibangun dari kumpulan data pelatihan dan yang divalidasi secara internal untuk menghasilkan nilai respon dari nilai prediktor[17]. Metode *Random Forest* (RF) merupakan metode yang dapat meningkatkan hasil akurasi, karena dalam membangkitkan simpul anak untuk setiap node dilakukan secara acak. Metode ini digunakan untuk membangun pohon keputusan yang terdiri dari *root node*, *internal node*, dan *leaf node* dengan mengambil atribut secara acak sesuai ketentuan yang diberlakukan.

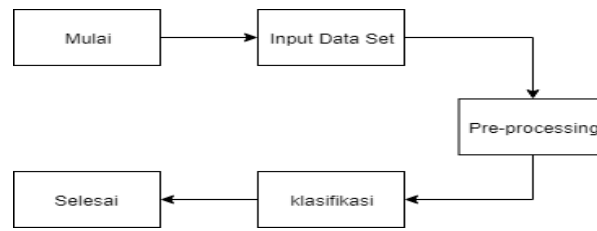


Gambar 2.1 Proses Klasifikasi Random Forest

3. Pembahasan

3.1 Gambaran Umum Sistem

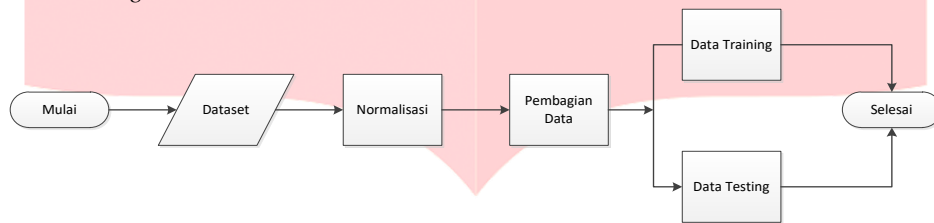
Berikut ini merupakan gambaran umum sistem dari aplikasi prediksi penentuan masuk SMP Negeri Kota Cimahi menggunakan metode *Random Forest*.



Gambar 3.1 Diagram Blok Sistem

Pada Gambar 3.1, berikut adalah penjelasan secara rinci mengenai proses pada aplikasi tersebut:

1. *Input Data Set*
Data yang akan digunakan sebagai data set yaitu data calon siswa yang akan melanjutkan jenjang pendidikan ke SMP Negeri Kota Cimahi.
2. *Pre-Processing*



Gambar 3.2 Flowchart Pre-processing

Proses *Pre-Processing* akan dilakukan secara manual dengan tujuan untuk membersihkan data set (*data cleaning*) yang kemudian akan di transformasikan langsung ke dalam bentuk code.

3. *Klasifikasi*
Setelah melakukan *data cleaning*, selanjutnya akan dilakukan klasifikasi menggunakan metode *random forest*. Data uji (*testing*) akan diuji dengan menggunakan data latih (*training*) yang sudah ada berdasarkan atribut – atributnya dan probabilitas yang didapatkan dari hasil pembelajaran.
4. *Hasil / Output*
Hasil dari semua proses pada aplikasi ini adalah prediksi terhadap objek yang diteliti dalam bentuk persentase.

3.2 Perancangan Sistem

Pada tahap perancangan sistem ini, akan dijelaskan lebih detail mengenai tahapan untuk *pre-processing*, *data training* dan *data testing* pada sistem prediksi masuk SMP Negeri Kota Cimahi menggunakan metode *Random Forest*. Adapun penjelasan tahapan untuk *data training* dan *data testing* adalah sebagai berikut:

3.2.1 Pre-Processing

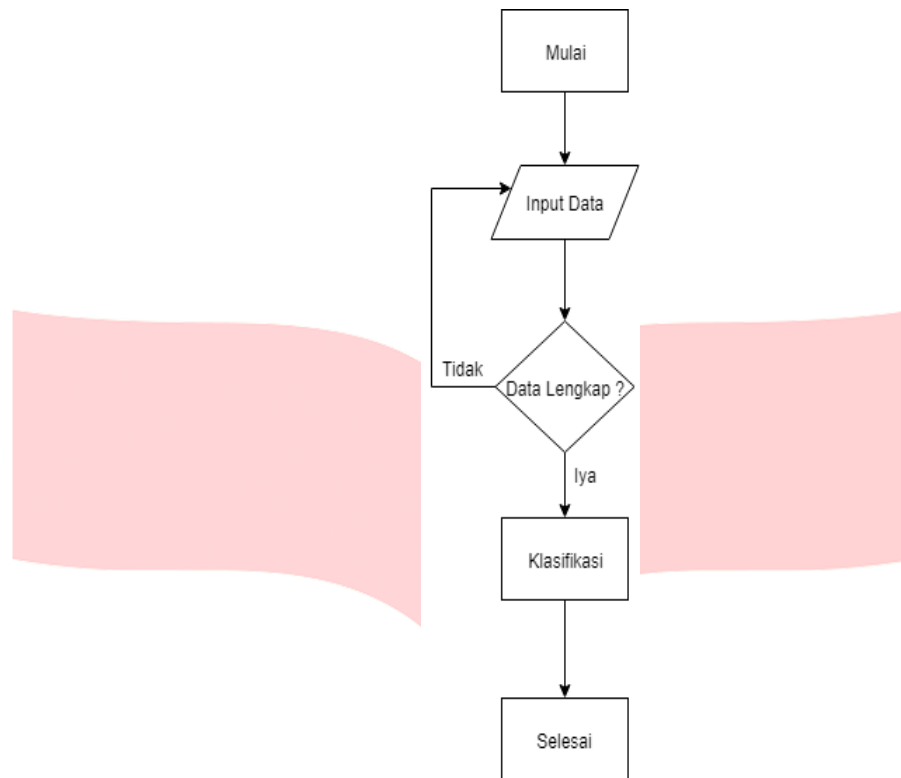
Tahapan *pre-processing* merupakan tahapan yang akan digunakan untuk membersihkan *data set* (*data cleaning*), dimana data mentah yang akan akan di normalisasi menjadi beberapa variabel yang sudah ditentukan untuk dijadikan data latih pada aplikasi ini. Berikut adalah proses untuk *pre-processing*.

3.2.2 Tahapan Data Training

Tahapan *data training* merupakan tahapan yang akan digunakan untuk pengenalan pola, *data training* sendiri yang akan digunakan adalah data seluruh SMP Negeri Kota Cimahi dalam kurun waktu 2 tahun terakhir ini (Tahun 2017 dan 2018). Atribut yang akan digunakan meliputi asal sekolah, domisili, nilai UN, lokasi rumah dan jalur prestasi siswa tersebut.

3.2.3 Tahapan Data Testing

Tahapan *data testing* merupakan tahapan yang akan digunakan untuk menguji pola yang telah didapatkan dari *data training*, *data testing* yang akan digunakan adalah data siswa yang akan mendaftarkan diri ke SMP Negeri Kota Cimahi. Data siswa yang digunakan adalah asal sekolah, domisili, nilai UN, lokasi rumah dan jalur prestasi siswa tersebut. Berikut merupakan *flowchart* untuk tahapan *data testing*.



Gambar 3.3 Flowchat Tahapan Data Testing

4. Pengujian

4.1 Metode Pengujian

Metode pengujian akan dilakukan menggunakan metode split *data testing* dan *data train* yang berbeda, dengan menggunakan *tools data mining* yaitu *WEKA*. Hasil yang didapatkan akan dibagi menjadi beberapa bagian nilai untuk memudahkan melakukan analisis, beberapa bagian nilai yang di dapatkan yaitu *correctly classified instances* dan *incorrectly classified instances* yang akan direpresentasikan dalam bentuk persentase dan *kappa statistic*. Parameter yang akan dijadikan sebagai parameter untuk pengevaluasian data yaitu *mean absolute error*, *root mean squared error*, *relative squared error* dan *root relative squared error*.

4.2 Pengujian Performansi

Pengujian perfomansi berdasarkan parameter yang telah ditentukan akan dilakukan pada data uji, berikut adalah data yang akan digunakan :

4.2.1 Pengujian Menggunakan Data Tahun 2018

Pengujian kedua akan dilakukan pada 11 data asli pada tahun 2018 sebagai masukan untuk penentuan hasil prediksi, berikut adalah tabel input dari pengujian kedua :

No	Siswa	Tahun	Domisili	Asal Sekolah	Nilai NUN	Jalur Prestasi	Jarak	Sekolah
1	L	2018	Dalam Kota	Dalam Kota	247.70	Tingkat Nasional Juara 1	distance_min > 1400 <= 1600	SMPN 1 Cimahi
2	M	2018	Dalam Kota	Luar Kota	276.90	Tidak Ada	distance_min > 1200 <= 1400	SMPN 2 Cimahi
3	N	2018	Dalam Kota	Dalam Kota	261.73	Tidak Ada	distance_min > 200 <= 400	SMPN 3 Cimahi
4	O	2018	Dalam Kota	Luar Kota	260.10	Tidak Ada	distance_min > 600 <= 800	SMPN 4 Cimahi
5	P	2018	Dalam Kota	Dalam Kota	248.18	Tidak Ada	distance_min > 1200 <= 1400	SMPN 5 Cimahi
6	Q	2018	Dalam Kota	Dalam Kota	229.12	Tingkat Nasional Juara 1	distance_min > 600 <= 800	SMPN 6 Cimahi
7	R	2018	Dalam Kota	Luar Kota	250.80	Tidak Ada	distance_min <= 200	SMPN 7 Cimahi
8	S	2018	Dalam Kota	Dalam Kota	258.85	Tidak Ada	distance_min > 1600 <= 1800	SMPN 8 Cimahi
9	T	2018	Dalam Kota	Dalam Kota	267.55	Tidak Ada	distance_min > 400 <= 600	SMPN 9 Cimahi
10	U	2018	Dalam Kota	Dalam Kota	285.10	Tidak Ada	distance_min > 1600 <= 1800	SMPN 10 Cimahi
11	V	2018	Dalam Kota	Dalam Kota	212.03	Tingkat Provinsi Juara 2	distance_min > 200 <= 400	SMPN 11 Cimahi

Gambar 4.1 Pengujian Kedua Menggunakan Data Pada Tahun 2018

No	Siswa	Hasil Prediksi	Persentase Masuk
		Random Forest	Random Forest
1	L	SMPN 1 Cimahi	90.78%
2	M	SMPN 2 Cimahi	100%
3	N	SMPN 3 Cimahi	100%
4	O	SMPN 4 Cimahi	100%
5	P	SMPN 5 Cimahi	100%
6	Q	SMPN 6 Cimahi	100%
7	R	SMPN 7 Cimahi	100%
8	S	SMPN 8 Cimahi	100%
9	T	SMPN 9 Cimahi	100%
10	U	SMPN 10 Cimahi	100%
11	V	SMPN 11 Cimahi	100%

Gambar 4.2 Hasil Pengujian Kedua

Pada Gambar 4.5 dapat dilihat hasil prediksi masuk sekolah pada metode *random forest* sesuai dengan data asli dan memiliki rata – rata persentase tingkat akurasi yaitu 99,16 %.

4.2.2 Pengujian Menggunakan Data Acak

Pengujian ketiga akan dilakukan pada 6 data input secara acak untuk penentuan hasil prediksi masuk sekolah, berikut adalah tabel input untuk pengujian ketiga :

No	Nilai NUN	Jalur Prestasi	Jarak		Hasil Prediksi	Persentase Masuk
			Latitude	Longitude	Random Forest	Random Forest
1	288.54	Tingkat Nasional Juara 2	-6.88514124	107.546204	SMPN 6 Cimahi	90%
2	263.49	Tingkat Provinsi Juara 3	-6.903446094	107.5319461	SMPN 9 Cimahi	60%
3	253.21	Tingkat Provinsi Juara 3	-6.875327478	107.5294801	SMPN 3 Cimahi	100%
4	255.34	Tidak Ada	-6.905638209	107.5430632	SMPN 9 Cimahi	50.00%
5	267.89	Tidak Ada	-6.906743915	107.5601944	SMPN 7 Cimahi	100%
6	248.19	Tingkat Internasional	-6.859656927	107.5398093	SMPN 5 Cimahi	80.00%

Gambar 4 3 Pengujian Ketiga Menggunakan Data Acak

Pada Gambar 4.6 dapat dilihat hasil prediksi masuk sekolah pada masing – masing metode dengan menggunakan *input* data secara acak, dimana terdapat beberapa hasil prediksi masuk sekolah tidak ada yang disebabkan oleh domisili tidak sesuai dengan *data set* yang sudah ada. Sedangkan untuk persentase masuk dengan menggunakan metode *random forest* memiliki nilai rata – rata tingkat akurasi sebesar 90%.

4.2.3 Hasil Pengujian Akhir Menggunakan Seluruh Data

Pengujian ini dilakukan pada 3946 data dengan 6 *attributes* yaitu domisili, asal sekolah, nilai UN, JP, Lokasi dan Sekolah. Pengujian dilakukan dengan menggunakan skema *weka classifiers* dengan *split data* 30 – 90 %.


```

=== Run information ===

Scheme:      weka.classifiers.trees.RandomForest -F 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:    Datasetbaru
Instances:   3946
Attributes:  4
            Nilai UN
            JP
            Lokasi
            Sekolah
Test mode:   split 30.0% train, remainder test

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.31 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.25 seconds

=== Summary ===

```

Correctly Classified Instances	2726	99.6966 %
Incorrectly Classified Instances	36	1.3034 %
Kappa statistic	0.9856	
Mean absolute error	0.01	
Root mean squared error	0.0479	
Relative absolute error	6.0585 %	
Root relative squared error	16.7146 %	
Total Number of Instances	2762	

Gambar 4.4 Hasil Prediksi split 30 % data train

Pada gambar 4.7 Dijelaskan informasi dari 30 % *data training* yang digunakan dan 70 % yang menjadi data test dengan *attributesnya* dan dijelaskan hasil dari *test* menggunakan skema *weka classifiers* dimana dapat dilihat pada hasilnya terdapat 2726 pohon prediksi dimana 2726 pohon prediksi dengan hasil akurasi kebenaran yaitu 98,6966 %.

5. Kesimpulan Dan Saran

5.1 Kesimpulan

Berdasarkan hasil uji coba sistem pada penelitian ini, didapatkan beberapa kesimpulan sebagai berikut :

1. Akurasi performansi yang diperoleh menggunakan metode Random Forest dalam 4946 data dari 11 SMPN di Kota Cimahi lebih dari **90% AKURAT**. Dari hasil akurasi tersebut dapat disimpulkan bahwa penggunaan metode Random Forest sangat baik.
2. Data Training yang semakin banyak maka akurasi performansi dari metode Random Forest semakin akurat.
3. Metode *random forest* memiliki tingkat keakuratan dan persentase masuk yang tinggi untuk setiap calon siswa di masing – masing sekolah sesuai dengan hasil prediksinya.

5.2 Saran

Saran untuk penelitian selanjutnya :

1. Disarankan mencoba Metode yang berbeda.
2. menggunakan data yang lebih banyak lagi dan juga lingkup yang luas.
3. Menambahkan atribut yang tidak numerik ke dalam sistem sehingga bisa ikut diolah oleh metode yang dipakai.

DAFTAR PUSTAKA

- [1] Kaur, M., Gulati, H., & Kundra, H. (2014). Data mining in Agriculture on crop price prediction: Techniques and Applications. *International Journal of Computer Applications*, 99(12), 1-2.
- [2] Luckyson Khaidem, Snehanshu Saha, Sudeepa Roy De, 2016. *Predicting the Direction of Stock Market Prices using Random Forest*.
- [3] Indrabayu, Nadjamuddin Harun, M. Saleh Pallu, Andani Achmad & Fikha C.L. 2012. Prediksi Curah Hujan Dengan Jaringan Saraf Tiruan. Vol. 6 PROSIDING 2012. pp. 1-8
- [4] Makridakis, S., Wheelright, S.C., dan McGee, V.E. 1992. Metode dan Aplikasi Peramalan - edisi ke-2, jilid I. Alih Bahasa : Andriyanto, U.S., dan Basith, A. Erlangga. Jakarta
- [5] Kaur, M., Gulati, H., & Kundra, H. (2014). *Data Mining in Agriculture on Crop Prediction: Techniques and Applications*. International Journal of Computer Applications, 1.

- [6] Kursini & Emha taufiq luthfi, 2009. Algoritma Data Mining. Yogyakarta :Penerbit Andi.
- [7] Girish K. Jha & Kanchan Sinha.,2013. Agricultural Price Forecasting Using Decision Tree: An Innovative Information Delivery System. *Agricultural Economics Research Review* Vol. 26 (No.2). pp. 229-239.
- [8] Silke Janitza, Jochen Kr uppa, Inke R. König Anne-Laure Boulesteix, "Over view of Random Forest Methodology and Practical Guidance with Emphasis on Computational Biology and Bioinformatics," *Technical Report Number 129, Department of Statistics, University of Munich*, 2012.
- [9] Friedman JH, Olshen RA, Stone CJ Breiman L, *Classification and Regression Trees*. New York: Chapman & Hall, 1993.
- [10] Breiman, Leo and Cutler, Adele. Random Forest. http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm, didownload pada tanggal 13 November 2006
- [11] Aggarwal, Charu.C., "Data Mining: The Textbook", Springer, IBM T.J Watson Research Center, New York.
- [12] Mulyadi, "Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penerima Beasiswa Prestasi," *Jurnal Sistem Informasi STMIK Antar Bangsa* Vol. 5, No. 2, Agustus 2016.
- [13] Bustami, "Penerapan Algoritma Naïve Bayes Untuk Mengklasifikasi Data Nasabah Asuransi," *Jurnal Informatika* Vol. 8, No. 1, Januari 2014.